Using extreme value theory for the estimation of risk metrics for capacity adequacy assessment

Using extreme value theory for the estimation of risk metrics for capacity adequacy assessment

Amy L. Wilson and Stan Zachary

***Abstract*—This paper investigates the use of extreme value theory for modelling the distribution of demand-net-of-wind for capacity adequacy assessment. Extreme value theory approaches are well-established and mathematically justified methods for estimating the tails of a distribution and so are ideally suited for problems in capacity adequacy, where normally only the tails of the relevant distributions are significant. The extreme value theory peaks over threshold approach is applied directly to observations of demand-net-of-wind, meaning that no assumption is needed about the nature of any dependence between demand and wind.**

The methodology is tested on data from Great Britain and compared to two alternative approaches: use of the empirical distribution of demand-net-of-wind and use of a model which assumes independence between demand and wind. Extreme value theory is shown to produce broadly similar estimates of risk metrics to the use of the above empirical distribution but with smaller sampling uncertainty. Estimates of risk metrics differ when the approach assuming independence is used, especially when data across different historical years are pooled, suggesting that assuming independence might result in the over- or under-estimation of risk metrics.

***Index Terms*—Capacity adequacy, reliability, risk, extreme value theory.**

I. Introduction

Wind generation plays an increasingly important role in the global supply of electricity. However, in the context of capacity assessment or security-of-supply the contribution of this wind generation can be difficult to quantify (see [1] and [2] for a survey of current approaches). This difficulty arises because, for capacity assessment, what matters most is the contribution of wind at the times of very peak demand when the system is typically under most stress. For example, in Great Britain (GB) this peak demand usually occurs in the early evening in winter when the weather is extremely cold. Because peak demand events are rare, and may scarcely occur at all in years with milder weather, there is relatively little data with which to make an accurate assessment of what the wind is doing at such times. Moreover, demand patterns are known to change through time, limiting the number of years of data suitable for estimating the relevant demand-wind relationship.

A particular concern is that the cold winter weather associated with the highest electricity demands may be associated with large-scale weather systems that lead to low wind conditions. There is some debate in the literature on this issue (see e.g. [3] and [4] and the references therein) but there is certainly insufficient evidence to suggest that there is no association at all between demand and wind (either positive or negative correlation). A failure to account for any reduction in wind at times of high demand would lead to overestimation of the contribution of wind to capacity adequacy.

The objective of a capacity adequacy study is to assess the risk of insufficient electricity generation to meet demand for some future year or season of interest. (Here by "season" is meant the peak demand season within each year, for example the winter months within GB.) Typically this is achieved using a risk metric such as *loss of load expectation* (LoLE) or *expected energy unserved* (EEU) ([5], [6]). Such expected value risk metrics may be completely defined in terms of a *non-sequential* or *snapshot* model which integrates over the course of the future season the joint sequential distribution of relevant variables such as demand, wind generation and conventional generation—without regard to the temporal structure of this joint distribution within the season (which is not necessary for the definition of such expected value risk metrics). Alternatively a non-sequential model may be viewed as that defining the distribution of the above variables at a uniformly randomly sampled point in time during the season under study (see [7] for further details).

The basic non-sequential probability model is well established, and used in Great Britain, the USA, and elsewhere (e.g. [8], [5]). The model consists of a specification of the joint distribution as described above of the random variables X, W and D which represent respectively *conventional generation*, *wind generation* and *demand*. Then the random variable

$$Z = X + W - D \tag{1}$$

models the corresponding excess of supply over demand. The model (1) has two major components:

(a) the (non-sequential) distribution of *demand-net-of-wind* $D - W$, which requires estimation from data;
(b) the (non-sequential) distribution of *conventional generation* X, which is usually given by a fully specified probabilistic model.

The variables $D - W$ and X are assumed probabilistically independent, so that the distribution of their difference Z is obtained by convolution. It is this distribution of *the supply-demand balance* Z which is the primary output of the non-sequential model, and from which the above risk metrics LoLE and EEU, and other statistics of interest, are calculated. For

The authors acknowledge support from EPSRC grants EP/N030028/1 and EP/K036211/1 as well as from National Grid and the School of Mathematics at the University of Edinburgh.

A.L. Wilson and S. Zachary are with the School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD, Scotland, UK (email: Amy.L.Wilson@ed.ac.uk, s.zachary@gmail.com).

more details of the underlying model see [9], [7], [10], [11]. In particular we have

$$\text{LoLE} = nP(Z < 0), \quad (2)$$

$$\text{EEU} = nE(\max(-Z, 0)) = n\int_{-\infty}^{0} \mathrm{d}z P(Z < z) \quad (3)$$

where n is the number of hours in the future season under study.

A standard approach for modelling the non-sequential distribution of conventional generation X is to use independent two-state models for individual generators and then to use convolution to obtain the distribution of the total available generation [12]. This paper is instead primarily concerned with the estimation of the distribution of demand-net-of-wind $D-W$. For the 'future' season to be studied, this distribution is typically estimated from a dataset of hourly historical paired observations of (demand, wind speed) made in earlier seasons. Each such historical season of demand observations is 'forward-mapped' by re-scaling to the future season under study. Forward-mapped wind generation observations for the future season are obtained by combining the historical wind speed measurements over a geographic grid with a physical model for the capacities, locations and power curves of the installed wind generation in the future season. For each historical season in the dataset this process yields a forward-mapped hourly trace (d_t, w_t) for $t \in \{1, \ldots, n\}$ of aggregate demand and wind generation for the future season under study. Estimates of risk metrics can be based on data from individual historic seasons or based on data pooled over multiple historic seasons.

One approach for estimating the distribution of $D-W$ is simply to use the empirical distribution of the forward-mapped observations $d_t - w_t$ (this is sometimes known as *hindcast*) ([7], [1]). This approach makes no assumption about the relationship between wind generation and demand. However, as we discuss above and illustrate in Section II, there is typically very little data available for estimating the part of the distribution of $D-W$ that is relevant for capacity adequacy, i.e. the far right tail. Further, as illustrated in Figure 2, the presence or absence of a single observation may, in the hindcast approach, have a very considerable influence on the estimated values of risk metrics such as LoLE and EEU. There is thus considerable concern as to the reliability of such estimates obtained using the hindcast approach.

An alternative approach to the estimation of the distribution of $D-W$ is to estimate the distribution of demand D and also the distribution of wind W conditional on demand D. The simplest possibility here is to assume that demand and wind are *independent*, in which case the distribution of wind may be estimated using all wind observations, not just those relatively few corresponding to times of high demand. However, as described above, there is concern that in GB, for example, the assumption of independence may not hold in practice. A more general approach is to develop a parametric (or smoothed) model for wind generation conditional on demand—see, e.g. [13].

This paper proposes a method for estimating the distribution of $D-W$ using statistical extreme value theory (EVT) [14]. EVT is a well-established methodology for making inference about the extremes of distributions and is therefore well-suited to problems in capacity adequacy, where, as discussed above and further in Section II, interest is in the extreme right tail of the distribution of $D-W$. EVT is based on asymptotic theory for the tails of distributions which permits appropriate smoothing and, if necessary, extrapolation of empirical data. As with the hindcast approach, EVT uses directly the empirical observations $d_t - w_t$ of $D-W$, without the need for any assumptions about the statistical relationship between demand D and wind generation W. The advantage of EVT is that smoothing the empirical data reduces the influence of the very small number of observations at the extremes of high demand and low wind. If appropriate, information about the shape of the far right tail of the distribution of $D-W$ can be inferred from observations that are close to the tail but not as extreme. A further advantage of EVT is that it does not make the assumption implicit in hindcast that there is no possibility of the demand-net-of-wind in the future year or season under study being more extreme than that observed historically. In [15], EVT is also used for capacity adequacy assessment, but in combination with quantile regression models that incorporate seasonal time effects. The advantage of the present approach is that it focuses on the use of EVT within the non-sequential model (which, as remarked above, is sufficient for the risk metrics considered) and avoids the considerable complications and distortions that arise in fully accounting for seasonal effects–which arise on multiple time scales (e.g. daily, weekly, yearly).

The present methodology was developed in response to concerns from the GB transmission system operator (National Grid) and was used to refine the estimates of risk metrics in the GB capacity adequacy study (see [8]). The GB system is therefore used as an exemplar. While the GB system is winter-peaking, this is not a necessary condition for the methodology to be appropriate—the same techniques are applicable for summer-peaking systems.

The rest of this paper is divided as follows: Section II describes the data used for the analysis of the GB system, Section III describes how EVT should be applied to model the supply demand balance Z, and Section IV presents and discusses results for the GB system.

II. Data

As described in the Introduction, the distribution of $D-W$ is estimated from hourly forward-mapped observations (d_t, w_t), $t \in \{1, \ldots, n\}$ of D and W. The season under study is the winter season of 2014–15, where winter is defined as the 21 weeks from the last Sunday in October. In GB, the risk of a shortfall at other times of year is negligible. The data described in [16] and [17] are used for the analysis. These data are also used in [13] and consist of:

- Hourly historical measurements of aggregate GB demand for the seven winter seasons from 2007–08 to 2013–14.[1]

[1] In GB, small (embedded) generators are not required to report their output to the transmission system operator so these demand measurements consist of the demand measured by the transmission system operator plus an estimate of the output of the embedded generators.

Measurements for earlier seasons are available but these were not thought to be representative of current demand patterns. These aggregate demand measurements are forward-mapped to the 2014–15 winter season by multiplying by an appropriate re-scaling factor for each of the seven seasons of historical data. These re-scaling factors were determined by calculating the 90% quantile of the daily maximum demands in each of the winter seasons from 1991–92 to 2013–14. A Lowess curve [18] was fitted to these 90% quantiles to obtain a smoothed curve estimating the variation in underlying demand over time. The re-scaling factor for the x-th season was then set to the ratio of the fitted value of the Lowess curve for the 2013–14 season (see below) to the fitted value for the x-th season.

This method of re-scaling historical demands by multiplying by some re-scaling factor is used in the GB capacity assessment study, where the re-scaling factor (calculated using a different methodology to that described above) is known as the Average Cold Spell (ACS) peak [19]. By re-scaling, the aim is to adjust historical demands for general trends, such as those due to changes in the economy, but to preserve any variation between (winter) seasons due to changes in weather. A Lowess curve is therefore appropriate because it smooths out any year-to-year fluctuations caused by random weather effects while still capturing long term trends in the re-scaling factor. By fitting the Lowess curve to the 90% quantiles rather than to the means the re-scaling is focused on the times of high demand that are of most interest in capacity assessment.

Note that the historical demands are re-scaled to the 2013–14 season, but the 'future' season under study is 2014–15. This mismatch is a result of data availability—the wind data related to January 2015, as described below, but the latest full year of demand data was 2013–14. We have therefore made an assumption that demand conditions in 2013–14 are similar to those in 2014–15. As the objective is to investigate methodology for assessing risk metrics this assumption has no effect on the conclusions drawn.

- Hourly aggregate GB wind generation 'observations' for the seven winter seasons from 2007–08 to 2013–14 forward-mapped to the 2014–15 winter season to pair with the forward-mapped demand observations described above. The wind generation observations were obtained by combining historical wind speed measurements (at the midpoint of each hour) with a model for the locations, capacities and power curves of the installed wind generation (approximately 14 GW of installed capacity) in January 2015. Aggregate wind generation for GB in each hour is then given by the sum of the wind power generated over all locations.

Figure 1 plots observations w_t of wind generation against corresponding demand d_t at the times of daily peak demand during the seven winter seasons comprising the dataset. A smoothed Lowess regression curve provides some evidence that the very highest demands may be associated with lower wind generation. Given the lack of data in the extreme region of interest, there is insufficient statistical evidence to decide the matter conclusively. However, the data do not justify any assumption of demand-wind independence.

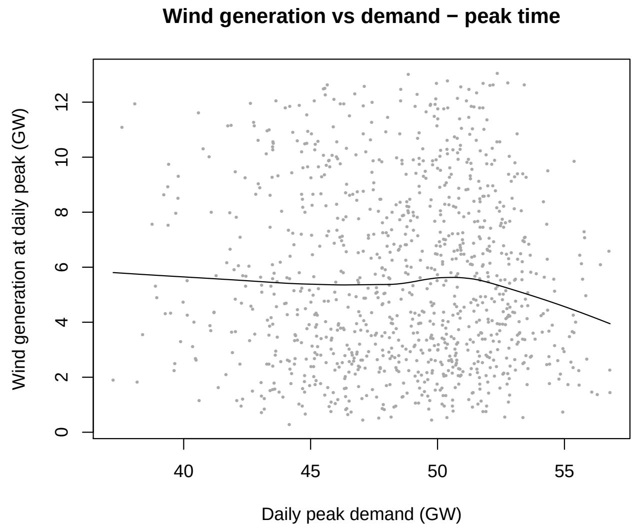

Fig. 1. Wind generation against demand at time of daily peak demand for seven winter seasons in GB. Overlaid is a Lowess curve showing the smoothed relationship between wind and demand.

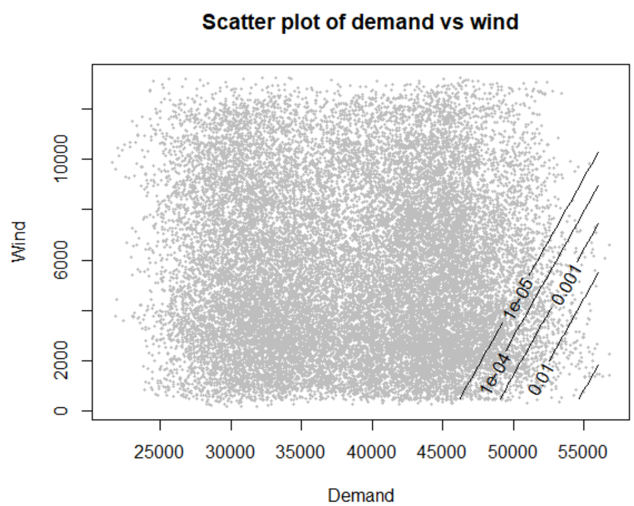

Fig. 2. Hourly wind generation (MW) against demand (MW) for seven winter seasons in GB. Contours show the contribution to LoLE of all points along the line.

Figure 2 plots all the hourly forward-mapped (demand, wind) data (d_t, w_t) for the seven winter seasons comprising the dataset. The contour lines separate points according to their values of $d_t - w_t$ and are such that the total contribution to the LoLE, of a point along the line in a hindcast calculation (and with the distribution of conventional generation X as described in Section IV-B) would be as indicated. Observe that the only points that make a significant contribution to this risk metric are indeed the very small number of observations in the lower right corner, i.e. in the extreme right tail of the distribution of $d_t - w_t$.

III. Methodology

A. Statistical model

We develop a statistical model for the marginal (non-sequential) distribution of demand-net-of-wind $(D - W)$ over the future season under study. The required result from EVT is that under appropriate conditions, which we discuss below, the tail of the distribution of $D - W$ is well-modelled by a generalised Pareto distribution (GPD). Specifically, the excesses

$$Y = D - W - u \tag{4}$$

of $D-W$ above any sufficiently large threshold u, conditional on $D - W > u$, have a distribution function H given approximately by

$$H(y) = P(Y \leq y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi}, \tag{5}$$

for y such that $y > 0$ and $1 + \xi y/\sigma_u > 0$ (see Chapter 4 of [14]). Here the shape parameter ξ is independent of the threshold u (for all u sufficiently large that the approximation (5) holds) and may be positive or negative, corresponding to whether the distribution of $D-W$ is heavy- or light-tailed. The parameter σ_u, which may be thought of as a scale parameter, depends on the threshold choice u, and increases linearly with it at rate ξ. The case $\xi = 0$ corresponds to $D - W$ having an exponential tail.

Once an appropriate threshold u is determined, and the parameters ξ and σ_u of the GPD estimated, the full distribution of demand-net-of-wind $D - W$ is given by its tail function

$$\begin{aligned} P(D-W > v) =& P(D-W > u)P(D-W > v \mid D-W > u) \\ =& P(D-W > u)(1 - H(v-u)), \quad v \geq u, \end{aligned} \tag{6}$$

for values of $D - W$ (here denoted by v) in excess of the threshold u. Here the probability $P(D-W > u)$ that $D-W$ exceeds the chosen threshold u is is taken to be its empirically observed estimate, i.e. the fraction of the observations of $D-W$ which exceed the threshold u. For values v of $D - W$ below the threshold u the probability $P(D-W > v)$ is again taken to be its empirically observed estimate.

Thus, for values of v below the threshold u, the estimates of the probability $P(D - W > v)$ will be the same for the EVT and the hindcast approaches, namely the empirically observed fraction of observations exceeding v. Where the use of EVT differs from the hindcast approach is that, for v above the threshold u, the empirical estimate of $P(D - W > v)$ is replaced by the smooth function given by (5) and (6). The effect of this smoothing is to reduce the influence of the very small number of observations in the extreme tail. This is because, for large v (perhaps considerably greater than the threshold u) the probability $P(D - W > v)$ is estimated by suitably smoothing the distribution of all the observations in excess of the threshold u, and not simply by the very small proportion of observations which may actually exceed v. (Clearly, for very large v, the empirical estimate is very sensitive to the precise number of observations in excess of v.) In particular, the EVT approach allows us to estimate $P(D-W > v)$ for values of v in excess of the largest observed value $d_t - w_t$ of $D - W$.

The result given in (5) is an asymptotic one. It assumes that the process of demand-net-of-wind has a sufficient degree of long-run stationarity—despite the existence of shorter-run seasonal variations—for the marginal distribution (represented above by the random variable $D - W$) to be meaningful. In addition there is an assumption of some mild regularity conditions and an absence of long-range dependence. (Further details are given in [14].) As historic years of data have been rescaled to the future year under study, we expect these assumptions to be reasonable.

The quality of fit by a GPD nevertheless requires empirical testing. Empirical methods are also required in the selection of an appropriate threshold u—one possibility is the examination of robustness of estimated parameter values across a range of thresholds. As we demonstrate in Section IV the GPD fit in general works very well for the GB data, and the parameter values are indeed robust with respect to threshold variation within a reasonable range.

Note that although (demand, wind) is a bivariate process our ultimate interest is in the univariate demand-net-of-wind distribution. The use of multivariate EVT methods to model the bivariate distribution offers no further advantage here.

B. Uncertainty

The aim of a capacity adequacy study is to assess the risk of insufficient generating capacity to meet demand in the future season under study. The LoLE and EEU are expected value metrics in that they give the long-run expected values of loss of load and energy unserved respectively. However, both wind generation and demand are very dependent on the weather, which varies considerably from one (winter) season to the next. Thus, estimates of LoLE and EEU conditional on the weather in a given season also vary considerably. To fully understand the risks to the system it is therefore important to understand this weather-dependent variation. This issue has become more important as the proportion of total energy supplied by variable generation has increased, as there has been a corresponding increase in variation in loss of load duration and energy unserved from winter season to winter season [20]. This year-to-year variation is reflected in the variation in the estimates of LoLE and EEU based on the forward-mapped demand-net-of-wind traces associated with individual historical seasons in our dataset. We therefore calculate these estimates based on individual historical seasons. The long-run LoLE and EEU are then estimated as the means of the respective individual season-based estimates.

It is also necessary to understand the statistical sampling uncertainty associated with the long-run LoLE and EEU estimates. This arises because there is considerable variation in weather conditions between years, and the long run estimates are based on a sample of a finite number (seven) of years of data. In addition to the considerable variation observed above in the estimates of LoLE and EEU based on the individual historical seasons of (demand, wind) data, there are further, within each historical season, complex patterns of dependence in the hourly forward-mapped ‘observations’ $d_t - w_t$ of demand-net-of-wind, including considerable positive autocorrelation and shorter-term nonstationarity—the latter due to both diurnal and seasonal effects. Hence, in making the above uncertainty estimates, the best that can reasonably be done is to block the data according to historical season and to treat these season-long blocks as being independent of each other. Where, as described above, separate estimates of LoLE and EEU are made based on each historical season of data, then confidence intervals for the long-run LoLE and EEU are given by the confidence intervals for the means of the seven independent historical-season-based estimates of these quantities. Since the individual season-based estimates of LoLE and EEU are far from normally distributed over seasons, we use a bootstrapping approach [21] in which, for example, the seven historical-season based estimates of LoLE are sampled with replacement to obtain a sufficiently large number of bootstrap replications of the original set of seven estimates. The distribution of the means of these bootstrap datasets mirrors that of the overall (sample) mean of the original seven season-based estimates, and so, in the usual bootstrap approach, the quantiles of this distribution may used to give confidence intervals for the 'true' long-run LoLE. These confidence intervals for the long-run LoLE and, similarly, the long-run EEU are arguably a little too narrow in that in each case the bootstrap approach effectively treats the extremes of the seven historical-season based estimates as representing the extremes of what may happen in any given season. However, the purpose of this paper is to investigate the use of EVT for estimating capacity adequacy metrics and—at least under the assumption of independence between seasons—the above bootstrap approach is sufficient to give a reasonably good approximation to the sampling uncertainty associated with the long-run estimates of LoLE and EEU.

As a comparison we also compute long-run LoLE and EEU estimates by combining the seven historical seasons of (demand, wind) data and obtaining pooled estimates of these quantities (i.e. using the full seven-year dataset to compute long-run metrics rather than taking the mean of the metrics corresponding to individual years). Confidence intervals may still be obtained by using block bootstrapping [22] in which the entire dataset is re-sampled in season-long blocks (assumed independent) to obtain a sufficiently large number of bootstrap replications of the entire dataset. Bootstrap estimates of LoLE and EEU are then obtained for each of these replications, and confidence intervals obtained as usual. Note that for the hindcast approach these two methods for obtaining confidence intervals for the long-run LoLE and EEU estimates will yield the same result. This is because the hindcast approach does not smooth between years when data are pooled.

IV. Results

In this section the model described in Section III is fitted to the GB data described in Section II and the LoLE and EEU are estimated using this model. The model is fitted using the `ismev` package [23] and the R computing language [24].

A. Model fitting and validation

To fit the model (6) to the demand-net-of-wind data it is first necessary to choose a threshold u. As results are to be calculated by conditioning separately on each forward-mapped historical season, different thresholds are used for each such season. As described in Section III, the distribution of demand-net-of-wind, for any given forward-mapped historical season, is modelled by a generalised Pareto distribution above the threshold u and by its empirical distribution below the threshold u. The threshold u must be sufficiently large that the required results from extreme value theory hold, but a lower threshold means that more data can be used for parameter estimation. The aim is therefore to use the lowest satisfactory threshold. Following [14] we test a range of values for the threshold u, fit the generalised Pareto distribution (5) for each u and assess the estimated values of the shape parameter ξ and scale parameter σ_u. For this purpose, the latter is transformed to $\sigma^* = \sigma_u - u\xi$ to remove the dependence of σ_u on the threshold u: we require to choose a u such that for all $u' > u$, there is little variation in the estimated values of ξ and σ^*.

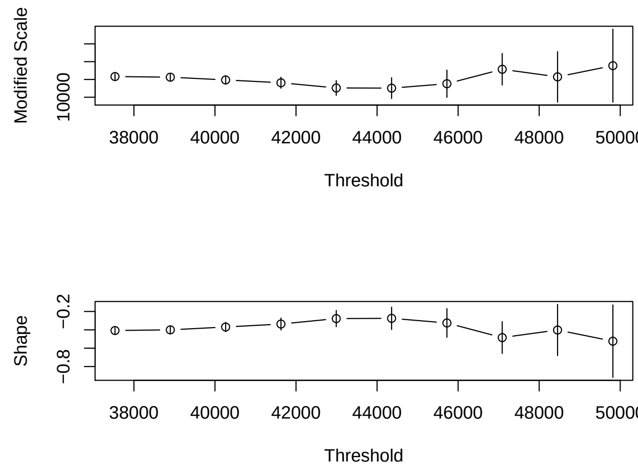

Fig. 3. Estimated values of ξ (bottom) and σ^* (top) for different values of the threshold u for the first season in the dataset. The bars show the uncertainties in the estimates.

Figure 3 shows the estimated values of ξ and σ^* for the first season of data, with thresholds u ranging from around 38GW to 50GW (approximately the 99.5% quantile). The uncertainties in the estimates of ξ and σ^* are clearly increasing as the threshold u increases as less data above the threshold is available for parameter estimation. These uncertainty estimates should be regarded as rough approximations, because their calculation treats the data within a given season as consisting of independent identically distributed observations, whereas, as previously remarked, there is actually some serial correlation structure within the data. Nevertheless, a threshold u of around 45GW appears to be reasonable. For thresholds less than 45GW there seems to be a trend in both parameters. For thresholds above 45GW the uncertainty intervals overlap to such an extent that there is no evidence to suggest that the parameter estimates are changing. For the first season in the dataset, 45GW corresponds approximately to the 95% quantile of the forward-mapped demand-net-of-wind data for that season. Repeating the analysis described above for each of the later seasons in the dataset leads to a similar conclusion—that in each case the 95% quantile is an appropriate choice of threshold. To further check the effect of threshold choice, the LoLE and EEU were estimated using thresholds corresponding to the 90%, 95% and 98% quantiles of the forward-mapped demand-net-of-wind data associated with each historical season. These results are discussed later along with further model validation to check that the fitted model is consistent with the data. The values of the 95% thresholds used for the analysis for each historical season of data, and also for a pooled analysis which combines the data over historical seasons are given in Table I.

GW	07-08	08-09	09-10	10-11	11-12	12-13	13-14	All
95%	45.28	45.38	46.57	47.42	44.88	46.88	43.64	43.89
σ_u	2.85	2.57	2.07	2.92	2.51	2.38	2.57	2.51
ξ	-0.32	-0.30	-0.22	-0.28	-0.24	-0.24	-0.38	-0.21

TABLE I

THRESHOLDS u (GW) FOR EACH SEASON IN THE DATASET (THE 95% QUANTILES) AND THE ESTIMATED SHAPE (ξ) AND SCALE (σ_u) PARAMETERS CORRESPONDING TO THIS 95% THRESHOLD. VALUES ARE LISTED FOR INDIVIDUAL SEASONS AND FOR THE FULL DATASET.

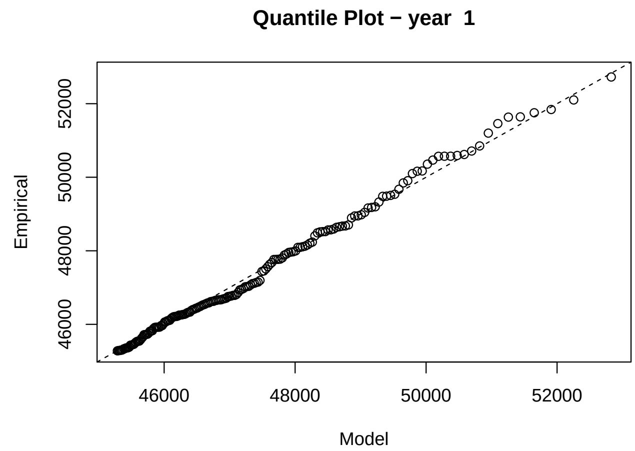

Fig. 4. Quantile-quantile plot showing the fit of model (6) to the extreme tail of demand-net-of-wind for the first season in the dataset.

Given the thresholds in Table I, maximum likelihood can be used to estimate the parameters ξ and σ_u in model (5) (using the `ismev` package). The parameter estimates are given in Table I for the threshold corresponding to the 95% quantile of the distribution of demand-net-of-wind. As shown, the estimates of ξ and σ_u are reasonably consistent from season to season, although the thresholds are variable. This year-to-year variability in the threshold suggests that a pooled analysis may not be entirely appropriate as datapoints that are extreme in one year may not be in another.

The fitted model can be validated by comparison to the observed data. Figure 4 is a quantile-quantile plot of the tail of the demand-net-of-wind data associated with the first historical season (i.e. the demand-net-of-wind data over the EVT threshold for that season) against the corresponding fitted model (6). The observed data are shown as circles. If the data followed the model (6) exactly they would lie on the dotted diagonal line shown. As can be seen, the data do very closely follow the fitted model. Quantile-quantile plots for the other historical seasons showed a similarly good fit.

B. Estimation of LoLE and EEU

It follows from (2) and (3) that estimation of LoLE and EEU is based on estimation of the distribution of the (non-sequential) supply-demand balance Z which is given by (1) and is the convolution of the corresponding distributions of demand-net-of-wind $D - W$ and conventional generation X. The fitted distribution of $D - W$ associated with any forward-mapped historical season of (demand, wind) data is entirely described by the tail function given in (6) above the threshold u and by the empirical distribution of the associated observations $d_t - w_t$ below that threshold. The distribution of conventional generation X is formed as described in [12]. Estimates of the capacities and availability probabilities of the conventional generators on the system in the future season of interest were obtained from National Grid. Random errors were added to the capacities to protect the sensitivity of the data. As such, the results presented should be seen as broadly representative of the GB system but do not provide accurate estimates of the risk in that system. Each generator is assumed to provide full capacity with its availability probability and otherwise to provide zero capacity. The generators are assumed to be independently available, and so the convolution of their individual two-state distributions gives the distribution of the total available conventional generation X. The distribution of Z—and so also estimates of LoLE and EEU—for the future season under study, based on any given historical season of (demand, wind) data, are then obtained as described above. The convolution of the distribution of X with that of $D - W$ is obtained by the discretisation of the latter in 1 MW bins.

Estimates of LoLE and EEU conditional on each historical season of (demand, wind) data, and for the three different choices of threshold u as above, are given in Tables II and III. The results are broadly similar for all three thresholds, suggesting that these estimates are not sensitive to the precise choice of threshold.

In Tables II and III these estimates can be compared to those obtained using two alternative approaches: hindcast and a model which assumes independence between demand and wind. The hindcast approach estimates the probability $P(D - W > v)$ by the empirical proportion of the observations $d_t - w_t$ which are greater than v. The model assuming independence fits separate empirical distributions to the demand and wind data for each historical season. The distributions for wind and demand are then convoluted to obtain the distribution of $D - W$. As shown in Tables II and III, the results obtained using the hindcast approach are similar to those obtained using EVT, especially to those EVT results obtained using a 98% threshold. (The latter observation is unsurprising since, as the threshold u is increased, the EVT analysis becomes closer to the hindcast.) However, using a lower threshold provides a greater degree of smoothing, inferring more information from less extreme data, and thereby providing results which are more robust to small changes in the data. The results obtained using the assumption of wind-demand independence are similar to those obtained using the EVT and hindcast approaches for some seasons of historical data and significantly different for others (e.g. 2008–09, 2009–10, 2012–13). For the 2008–09 and 2009–10 historical data the risk levels obtained using the independence assumption are higher, while for the 2012–13 historical data the risk level obtained using the independence assumption is lower. Since it is the above independence assumption which is suspect here, these results highlight the dangers involved in making it.

Season	EVT 90%	EVT 95%	EVT 98%	Hindcast	Ind
07-08	2.86	2.82	3.00	3.07	2.74
08-09	2.21	2.22	2.25	2.29	3.05
09-10	4.43	4.02	3.90	3.85	5.45
10-11	16.33	16.77	17.63	17.60	17.81
11-12	2.17	1.92	1.92	1.95	1.56
12-13	7.87	7.69	7.57	7.97	5.64
13-14	0.16	0.15	0.15	0.17	0.26
Mean	5.15	5.08	5.20	5.27	5.22
CI	(2.02,9.25)	(1.92,9.37)	(1.95,9.71)	(1.97,9.79)	(2.01,9.70)

TABLE II

LoLE (hours per season) conditional on the demand-wind profile of each season in the dataset, estimated using: extreme value theory (with thresholds of the 90% quantile, 95% quantile and 98% quantile), hindcast and the independence model (Ind). Confidence intervals account for sampling uncertainty in demand and wind.

Season	EVT 90%	EVT 95%	EVT 98%	Hindcast	Ind
07-08	2.99	2.81	2.92	3.03	3.02
08-09	2.14	2.12	2.13	2.18	3.07
09-10	4.56	4.15	4.07	4.09	6.11
10-11	24.07	24.01	25.01	25.92	24.77
11-12	2.11	1.95	1.96	2.05	1.45
12-13	9.22	9.16	9.16	9.73	6.17
13-14	0.11	0.10	0.10	0.12	0.19
Mean	6.46	6.33	6.48	6.73	6.40
CI	(2.02,12.73)	(1.91,12.61)	(1.92,13.04)	(1.97,13.53)	(2.07,12.84)

TABLE III

EEU (GWh per season) conditional on the demand-wind profile of each season in the dataset, estimated using: extreme value theory (with thresholds of the 90% quantile, 95% quantile and 98% quantile), hindcast and the independence model (Ind). Confidence intervals account for sampling uncertainty in demand and wind.

The results in Tables II and III show substantial variability between seasons. The LoLE ranges from around 0.15 to 16.77 hy^{-1} and the EEU ranges from around 0.1 to 24.01 GWhy^{-1}. That these ranges are wide highlights the need to consider the variability of the risk level with weather conditions. Decision-makers might be very averse to an LoLE or EEU at the higher end of this range but happy with the overall mean. Note that the estimates of LoLE and EEU conditional on a given demand-net-of-wind profile still integrate over uncertainty in conventional generation (and hence are still expected values). The actual variation in loss-of-load and energy unserved from season to season will therefore be larger than the variation shown between seasons in Tables II and III.

Tables II and III also give estimates of the long-run LoLE and long-run EEU. These are the means of the estimates based on the individual historical seasons of demand-net-of-wind data. The 95% confidence intervals for these long-run estimates are calculated via bootstrapping as described in Section III, i.e. by regarding the estimates for the seven historical seasons as seven independent observations. The confidence intervals for the long run estimates therefore reflect uncertainty arising from the limited number of seasons of data (see also Section III-B). These confidence intervals are wide, ranging from around 2 to 10 hy^{-1} for LoLE and from 2 to 13 GWhy^{-1} for EEU, reflecting the considerable variability in the estimates based on individual seasons of data. Again, this variability demonstrates the importance to decision-makers of understanding this uncertainty. The widths of the confidence intervals increase slightly with increasing EVT threshold and are greatest for those based on the hindcast approach. This is to be expected—the smoothing provided by the EVT approach reduces variability because more data is used to estimate the far right tail of the supply-demand balance. The widths of the confidence intervals obtained under the assumption of demand-wind independence are comparable to those obtained using EVT but, as described above, the use of the independence assumption risks biasing the estimates of LoLE and EEU.

Table IV gives pooled estimates of LoLE and EEU obtained by fitting the above models to all seven historical seasons of data simultaneously (in contrast to fitting the models to each season individually). The EVT thresholds used are the corresponding quantiles of the demand-net-of-wind distribution for the full dataset. For the EVT and hindcast approaches, the long-run LoLE and EEU estimates are similar to those already obtained as the means of the individual season-based estimates and reported in Tables II and III. The corresponding 95% confidence intervals—obtained using block bootstrapping with season-long blocks—are unsurprisingly also similar to those already obtained, with the hindcast approach again giving the widest confidence intervals. For the model based on the assumption of demand-wind independence, the pooled estimates of LoLE and EEU are respectively 4.46 hy^{-1} and 5.30 GWhy^{-1}, in contrast to the earlier individual season-based estimates of 5.22 hy^{-1} and 6.40 GWhy^{-1}. For this independence model the confidence intervals obtained via the pooled analysis are smaller than those obtained previously. These results suggest that the assumption of demand-wind independence may be more problematic when applied across multiple seasons of data. One possible reason is that long-term weather regimes may induce dependence between demand and wind generation aggregated over multiple seasons, while this dependence may largely disappear when conditioning on individual seasons (as in Tables II and III). These results suggest that if a demand-wind independence model is used, better results may be obtained by fitting the model separately to each season in the dataset and then obtaining risk metrics by averaging over these seasons.

	EVT - 90%	EVT - 95%	EVT - 98%	Hindcast	Ind
LoLE	5.36	5.26	5.10	5.27	4.46
95% CI	(1.96,9.46)	(1.93,9.33)	(1.93,9.52)	(1.97,9.79)	(1.79,8.66)
EEU	6.53	6.52	6.55	6.73	5.30
95% CI	(1.94,12.72)	(1.95,12.84)	(1.95,12.92)	(1.97,13.53)	(1.84,11.28)

TABLE IV

LoLE (hours per season) and EEU (GWh per season) estimated by fitting the models to the full seven season forward-mapped demand-net-of-wind dataset. Models used: extreme value theory (with threshold of the 90%, 95% and 98% quantiles), hindcast and the independence model (Ind).

V. Conclusion

This paper has investigated the use of extreme value theory (EVT) for modelling the distribution of demand-net-of-wind for capacity adequacy assessment. The main advantage of this approach is that EVT provides a mathematically-justified mechanism for estimating the extreme right tail of the distribution of demand-net-of-wind (corresponding to times of high demand and low wind); this is normally the only part of the distribution which is relevant for capacity adequacy. This smoothing involved in this estimation reduces the effect of outliers and small variations in the tail data when compared to use of the empirical distribution. A further advantage of this approach is that observations of demand-net-of-wind can be used directly, meaning that there is no need to make strong parametric assumptions about the underlying distributions of the demand and wind processes, or about the nature of the dependence between demand and wind.

The paper has also shown that typically estimates of risk metrics such as LoLE and EEU vary greatly according to the historical winter season of (demand, wind) data used in the estimation process, indicating a strong dependence on the prevailing weather in the winter season under study. This has two consequences: first, actual outcomes for these metrics in any given future season may be very different from estimated long-run averages; second, uncertainty estimation for these long-run averages can probably only be satisfactorily made by blocking data according to historical season and treating (demand, wind) regimes in distinct winter seasons as independent—observations within seasons may not be treated as independent of each other. The first consequence is further compounded because there are usually only a small number of relevant years of data for the estimation of risk metrics (seven in our example), meaning that it is unlikely that the full year-to-year variability has been captured in the dataset.

Acknowledgment

The authors would like to thank Iain Staffell for providing the wind and demand data and National Grid for providing data on conventional generation. They are further grateful to Chris Dent and to colleagues at National Grid for helpful comments and discussions.

References

[1] A. Keane, M. Milligan, C. J. Dent, B. Hasche, C. D'Annunzio, K. Dragoon, H. Holttinen, N. Samaan, L. Soder, and M. O'Malley, "Capacity value of wind power," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 564–572, 2011.

[2] M. Milligan, B. Frew, E. Ibanez, J. Kiviluoma, H. Holttinen, and L. Söder, "Capacity value assessments of wind power," *WIREs Energy Environ*, vol. 6, no. 1, pp. 1–15, 2017.

[3] D. J. Brayshaw, S. Zachary, and C. J. Dent, "Wind generations contribution to supporting peak electricity demand - meteorological insights," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 226, no. 1, pp. 44–50, 2012.

[4] H. E. Thornton, A. A. Scaife, B. J. Hoskins, and D. J. Brayshaw, "The relationship between wind power, electricity demand and winter weather patterns in Great Britain," *Environmental Research Letters*, vol. 12, no. 6, 2017.

[5] NERC, "Probabilistic adequacy and measures," https://www.nerc.com/comm/PC/Documents/2.d_Probabilistic_Adequacy_and_Measures_Report_Final.pdf, 2018, accessed: 2019-06-14.

[6] M. Amelin, "Comparison of capacity credit calculation methods for conventional power plants and wind power," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 685–691, 2009.

[7] S. Zachary and C. J. Dent, "Probability theory of capacity value of additional generation," *Proceedings of the Institution of Mechanical Engineers, part O: Journal of risk and reliability*, vol. 226, no. 1, pp. 33–43, 2012.

[8] National Grid, "National Grid Electricity Market Reform electricity capacity report," https://www.emrdeliverybody.com/Lists/Latest%20News/Attachments/47/Electricity%20Capacity%20Report%202016_Final_080716.pdf, 2016, accessed: 2019-07-30.

[9] A. L. Wilson, S. Zachary, E. Ibanez, M. Milligan, J. Dillon, E. Lannoye, A. Tuohy, and C. J. Dent, "Capacity adequacy and variable generation: Improved probabilistic methods for representing variable generation in resource adequacy assessment," *EPRI*, 2016.

[10] C. J. Dent and S. Zachary, "Estimation of joint distribution of demand and available renewables for generation adequacy assessment," 2014, preprint available at http://arxiv.org/abs/1412.1786.

[11] C. J. Dent, R. Sioshansi, J. Reinhart, A. L. Wilson, S. Zachary, M. Lynch, C. Bothwell, and C. Steele, "Capacity value of solar power: Report of the IEEE PES task force on capacity value of solar power," *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016.

[12] R. Billinton and R. N. Allan, *Reliability evaluation of large electric power systems*. Springer, 2012.

[13] A. L. Wilson, S. Zachary, and C. J. Dent, "Use of meteorological data for improved estimation of risk in capacity adequacy studies," *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018.

[14] S. Coles, *An introduction to statistical modeling of extreme values*. Springer, 2001.

[15] W. Gao and D. Gorinevsky, "Probabilistic balancing of grid with renewables and storage," *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018.

[16] I. Staffell and S. Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, vol. 114, pp. 1224–1239, 2016.

[17] ——, "The increasing impact of weather on electricity supply and demand," *Energy*, vol. 145, pp. 65–78, 2018.

[18] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.

[19] National Grid, "Average cold spell methodology," https://www.emrdeliverybody.com/Lists/Latest%20News/Attachments/189/SC4L12%20ACS%20Methodology.pdf, 2018, accessed: 2018-09-30.

[20] S. Sheehy, G. Edwards, C. J. Dent, B. Kazemtabrizi, M. Troffaes, and S. Tindemans, "Impact of high wind penetration on variability of unserved energy in power system adequacy," *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016.

[21] A. Davison and D. Hinkley, *Bootstrap methods and their application*. Cambridge University Press, 1997.

[22] D. Politis, "The impact of bootstrap methods on time series analysis," *Statistical Science*, vol. 18, no. 2, pp. 219–230, 2003.

[23] J. E. Heffernan and A. G. Stephenson., *ismev: An Introduction to Statistical Modeling of Extreme Values*, 2016, R package version 1.41. [Online]. Available: https://CRAN.R-project.org/package=ismev

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2018. [Online]. Available: https://www.R-project.org/

Amy L. Wilson is a Research Associate in the School of Mathematics, University of Edinburgh, U.K. She obtained her MMath degree in Mathematics from the University of Cambridge, U.K. and her PhD in Statistics from the University of Edinburgh. Her current research interests include the analysis of risk for capacity adequacy studies and quantification of uncertainty in large scale energy system models.

Stan Zachary is a Senior Researcher in the School of Mathematics, University of Edinburgh, U.K, and an Honorary Fellow of Heriot-Watt University, Edinburgh. He obtained his MMath degree in Mathematics from the University of Cambridge, U.K. and his PhD in Statistics from the University of Durham. His current research interests include the probabilistic, statistical and economic modelling of large energy systems.